Whois Misuse Study Webinar
17 December 2013 at 19:00 UTC
ICANN Transcription

Mary Wong: Welcome everyone.

In (deference) to those of you who dialed in early or on time, we are going to begin this webinar. And to kick us up, I'm going to handle the first to my colleague, Nathalie Peregrine. Nathalie, could you do the honors, please?

Nathalie Peregrine: Thank you, Mary. This is Nathalie. Just to (let you know) - you probably realize we're currently streaming audio in Adobe Connect Room. If you would like at the end to take part in the question and answer session over the audio, please don't hesitate to dial into the audio bridge, the passcode is whois. Otherwise, you're very welcome to type your questions into the chat. If you have any audio issues during the call, please don't hesitate to private message me and I'll be happy to assist. Thank you and back to you, Mary.

Mary Wong: Thank you, Nathalie, and again, my name is Mary Wong, I'm on the ICANN policy staff, and to follow up on what Nathalie said, for those of you who are on just the audio bridge or the phone bridge and not the Adobe Connect Room, if you would like to join us in the Adobe Connect Room, that might be helpful as we have slides up that might help you to follow along.

What we will do is we will start off with the presentation by our researchers and have a question and answer session at the end, at which point as Nathalie mentioned, your mics will be unmuted. As we are listening to the

webinar, should you have any comments or questions, you can also type them into the chat room in the Adobe Connect Room if you are there, and during question and answer, we will read out your question or you can raise your hand in Adobe Connect or speak on the phone bridge.

So I will just kick things off with a few introductory remarks. First, I would very much like to welcome our researchers (unintelligible) study, Dr. Nicolas Christin and Nektarios Leontiadis. As many of you know, this is one of a series of studies and surveys that have been commissioned by the GNSO Council in the last few years. This all followed from a resolution by the GNSO Council in 2007, who wanted to obtain objective, quantifiable on certain important aspects of the WHOIS system.

On this introductory slide, you see some of the other studies of surveys that were commissioned, who they were done by, and when those reports were released for public comment. This therefore means that this WHOIS misuse study by our CMU team, when done, will complete all of the GNSO's current projects on WHOIS. All these slides will be made available to you at the end of the webinar and published on the GNSO Council Web site, and there is a link to all the various studies and reports on these projects that I've just mentioned.

Like I said, we're very honored that we have our presenters here with us today. From CyLab at Carnegie Mellon University, there's some information here on this slide about CyLab, which is an extremely well known research and academic organization and as well as a Center of Academic Excellence in various respects with the NSA.

The team is led by Dr. Christin, who is an assistant research professor with CyLab as well as CMU's Department of Electrical and Computer Engineering, and he was assisted by Nektarios who I believe will be presenting the slides today. And today's webinar they will speak to the hypothesis, the

methodologies they use, the studies they did, as well as give an overview of some of the findings and their analyses.

We've included links to their bios so that you can look at those at your leisure. So welcome Nicolas, welcome Nektarios, and over to you.

Nektarios Leontiadis: Thank you Mary for the introduction. This is Nektarios and right next to me is Nicolas, and I'll start with this webinar and we will be presenting the, as you said, the main methodological aspects and (unintelligible) of the study we conducted her at CMU in response to ICANN's decision to pursue a set of five WHOIS studies characterizing the use and misuse of several aspects of the WHOIS service.

And the empirical findings of those studies intend to inform (unintelligible) decisions at (unintelligible) in terms of the future of WHOIS. The present study tries to empirically verify the hypothesis that public access to domain registrar information through WHOIS leads to a measurable degree of misuse. Furthermore, if this hypothesis is validated, this study aims at identifying the main forms and characteristics of WHOIS misuse and the effectiveness of anti-harvesting mechanisms in protecting registrar information.

The scope within which the study tries to answer those questions are defined by the terms of reference for WHOIS studies. More specifically we look at the top five generic top-level domains that in 2011 represented about 98% of the total registered domains. The same scope also applies to the other four WHOIS studies that ICANN has commissioned.

Methodologically, this work is divided into two main components. The first component is a descriptive study that uses a set of interviews and surveys to capture the details of past instances of WHOIS misuses. All surveys were done independently by CMU with explicit (unintelligible) of response privacy and anonymity.

The second component, the experimental study, tries to measure empirically and (systematically) the occurrence of WHOIS misuse and ideally corroborate the findings with the ones from the descriptive study. I will start with a discussion of the various components of the descriptive study and the associated key results and then I will continue with the experimental part.

As part of the descriptive study, we constructed three questionnaires, targeting three key groups. The intent was to capture different but complementary perspectives on the occurrence of WHOIS misuse. The first survey targeted registrars, asking them in what ways, if any, the information (unintelligible) WHOIS has been misused in the past. More importantly, we sought to find whether they could attribute this misuse to their information being listed in the WHOIS.

The second survey targeted registrars and registries that maintain its own files or the file of top-level domains, and the intent was to get aggregate data on the occurrence on WHOIS misuse and to identify the methods they use to protect their registrants' information. The third component was an expert survey targeting cybercrime researchers, law enforcement, and consumer and data protection organizations aiming at getting a more holistic perspective on online crime in general and more specifically on the occurrence and significance of WHOIS misuse.

Now starting with a registrant survey, we build a representative registrant sample based on a set of 6000 randomly-selected domains from the five top-level domains, which we then reduced into a microcosm of 2900 domains having the same TLD distribution as the population of domains in those five TLDs. Given the desire for this (unintelligible) significance, we said (unintelligible) target to 340 registrants.

Despite our repeated and (unintelligible) to reach this goal, we managed to collect 57 responses, which brings the margin of error up to 12.7%. This is

not as bad as it may seem initially as our intention is not to show that WHOIS misuse is beyond a certain rate but rather to a certain whether the misuse or (unintelligible) at a statistically significant level, and that is exactly what we discovered.

Based on the collected reports, WHOIS misuse occurs at a statistically significant level. Almost 44% of the registrants participating in our survey were affected by one or more types of WHOIS-attributed misuse. That is misuses that the registrants could directly associate with the exposure of their personal information in the WHOIS.

The participants reported three main types of misuse, personal address, email address, and phone number misuse. The first two represented by postal and email spam are the most prominent ones with about 30% of the participants reportedly being affected. Phone number misuse, for those, with about 12% of participants being affected. Other types of reported harmful acts either occurred at a very small scale like blackmail or the registrant could not associate the harmful acts with WHOIS misuse.

Now we'll move on to the second survey targeting registrars and registries. The registrars we consider for participation in the survey with 107 registrars associated with a domain in the registrant sample. We also included the four registries responsible for the five top-level domains within the scope of the study. Overall we got 22 responses from registrars and a single response from registry.

Moreover, the registrars and registries participating the survey opted to leave many questions in the survey unanswered. In this survey, we did not make of statistical significance but the responses represent the view of 22 of the 107 most popular registrars and of one top-level domain registry.

Moving on to the findings of the survey, registrars and registries reported email spam as the most prominently-reported incident of WHOIS-attributed

misuse. That is followed by (facing) attempts, postal spam, malicious software sent via email, attempts for identity theft, and various forms of blackmail.

Six participants reported being able to verify that they reported incidents where close by WHOIS misuse. While these results are based on the reports received at the registrars and registries, it is important to know how often they directly observe efforts of WHOIS harvesting and what do they do about it. 30% of the participants reportedly observed the terms of WHOIS harvesting but no participant claim that any (unintelligible) was successful.

Also, 57% reportedly implemented one or more WHOIS anti-harvesting method. These methods include IP blacklisting, (query rate) limiting on port 43, which is the designated WHOIS clearing port, (unintelligible) challenges, and private proxy registration services. I need to stress that once again that a big chunk of the participants did not offer any answer to this set of questions, so these findings are more indicative of the current situation rather than representative.

Before moving on to the third survey, I will talk briefly about a small experiment that we used to test the registrars and registries for the existence of WHOIS anti-harvesting techniques. We found that about 51%, the majority of registrars and registries, do not employ any port 43 rate limiting technique. The remaining portion uses a variety of (unintelligible) similar to the ones reported in the survey.

Now moving on to the expert survey, we built up on our contacts here at CMU and on ICANN's context to assemble a panel of security researchers and consumer and government organizations with special (unintelligible) would involve awareness of a specific incidence of WHOIS misuse.

As I mentioned earlier, the goal for the survey was to get a holistic perspective on online crime in general and more specifically on the

occurrence and significance of WHOIS misuse. In total, we recruited 101 participants, mainly security researchers and law enforcement agents. Moreover, while we invited participants from all geographic regions, we mainly have representation from the Americas and Europe.

Overall, the (unintelligible) details on 23 incidents of WHOIS misuse with almost half of them targeting directly the participants. Most cases involve spam email with marketing material or bills, but we also had a few reports on more sophisticated attacks. In about half of the reported cases, the victims did not take any protective measures after the attack while the other half reportedly attempted to avoid further misuse by deploy specific countermeasures like IP blocking.

In terms of the portion of the WHOIS information being misused, the email address has reportedly the highest occurrence with 70% of the cases followed by the registrar name and postal address in 26% of the cases. Finally, the phone number was misused in 17% of the reported cases.

And now we're going to move on to the experimental part of the study, and also by discussing the overall methodology. The goal of the experimental study was to measure empirically and systematically the occurrence of WHOIS misuse. To this end, we registered 400 domains across the five top-level domains using a representative sample of registrars.

Each of those domains was associated with an artificial registrant identity and the domain names were crafted in such a way as to be categorized in one of five categories of interest. Over a period of six months, we used this experimental platform to measure the occurrence of WHOIS misuse as it was reflected by the number of spam emails, postal mail spam, and voice spam collected as voicemail.

In selecting the registrars we used to register the experimental domains, we started with 107 registrars associated with the domains in the registrant

sample. And we picked 16 based on a set of criteria we set. A key criterion was the popularity of the registrar as it was reflected through the representative registrant sample.

In total we registered 25 domains per registrar, which essentially means five domains for each of the five top-level domains. Each of those five domains was selected from one of the five domain categories. These categories are the following. Strings of completely random letters and numbers, strings representing personal names in the form of first name (dash) last name, synthetic names composed by concatenating two random words from the English vocabulary, names representing businesses in ten professional categories often abused through facing by online criminals, and finally names from four professional categories that are not known to be targeted.

Moving on to the artificial registrant identities, as our audience would probably know, a registrant identity is composed the registrant's full name or organization, phone number, and postal and email addresses. So, each one of the 400 registrant identities have all those pieces of information. We constructed the registrar names by piecing together random combinations of first and last names.

Also, each identity had one public email address and a set of private email addresses. The public one was published only through WHOIS and it was in the form of contact.domainname.dld. The set of private email addresses was not listed anywhere. In terms of postal addresses, we reused across all identities the addresses of three PO boxes located in the US.

Our initial intention was to use a set of geographically diverse postal addresses different for each registrar and identity but we encountered numerous difficulties in this effort, and you may find more details on that in the report. In regards to phone numbers, we acquired 80 Skype numbers for the duration of the experiment and all incoming calls were sent directly to voicemail. I should mention that each phone number was reused between the

experimental domains registered at the same registrar and under the same top-level domain.

Now, having discussed the experimental methodology, I am moving on to the key findings for this experiment and I will start with a case of WHOIS attracted email address misuse. 95% of all emails we collected are the public email addresses was classified as SPAM and was directed to 71% of the experimental domains.

I need to stress that our methodology allows us to say with high certainty that all of these measured misuse easily attributed in fact to WHOIS. Moving on to the occurrence of WHOIS-attracted phone number misuse, we collected in total 674 voicemails and 39 of those were classified with high certainty as WHOIS-attributed spam. The Skype accounts receiving those calls were associated with 30% of the registered experimental domains, primarily under the .biz, .info, and .com DLDs followed by .net and .org.

It is not worthy that we faced a few challenges in classifying the voicemails. For example, it's not uncommon that any one of us may get an unsolicited phone call only to realize that in fact the other party called our number by accident. This means that we could not on any automated classification system but we had to manually classify each and every piece of received voicemail.

The third and final type of misuse we identified targeted the registrants' postal addresses. In total we collected four pieces of WHOIS-attributed postal spam and 34 pieces of postal spam that was not associated with WHOIS misuse. While this extent of postal address misuse does not allow for any meaningful statistical analysis, the mere occurrence of WHOIS-attributed postal spam suggests that registrants' postal addresses are in fact targeted and misused.

Now, correlating the findings of the experimental study with once we reported in the descriptive study, we find that through both routes we identified the

same major types of WHOIS misuse. These are email address, postal address, and phone number misuse. In comparing the rates of misuse, only the case of the phone number misuse, there's (much) between the measured and the reported rates. In the case of postal address misuse, we measured a much lower rate and we believe that this may be caused by the limitation I mentioned earlier.

Finally, the lower frequency of reported email address misuse may be ascribed to the difficulty the registrants may have in classifying email spam as WHOIS-attributed. As most of us received significant loads of spam email and the modern spam filters do a good job in keeping it out of our sight. We further performed a statistical analysis to identify significant correlations between the measured misuse and the characteristics of the experimental domains.

We considered the following possible contributing characteristics. The existence of WHOIS anti-harvesting measures, the top-level domain, the price we paid to get each of the experimental domains, the category of the domain name, and the domain registrars.

In terms of the observed phone number misuse, we found that the only factor that had a statistically significant contribution was that of the top-level domain. We found .biz and .info domains were subject to more WHOIS-attributed misuse while .org domains were subject to less misuse.

Considering now the observed email address misuse attributed to WHOIS, we found a number of statistically significant correlations. First, the lack of WHOIS anti-harvesting measures is linked to 2.3 times more misuse. Furthermore, looking at the top-level domain, we found that .biz domains are, again, subject to more misuse while .com, .net, and .org domains see fewer WHOIS-attributed spam emails.

Moreover, higher-price domains are correlated with less misuse and the same applies for domain names denoting a person name. And you may see in the table, we did not find any evidence to support a correlation between any type of misuse and the registrars. Now before moving to Q&A, I will briefly summarize the findings with the study.

We found statistical evidence through a combination of a descriptive and the experimental study, that public access to WHOIS leads to a measurable degree of misuse. 44% of the registrars participating in our survey have directly experienced WHOIS misuse, which most prominently affects their email addresses, their postal addresses, and the phone numbers published in the WHOIS.

Finally, the evidence suggests that WHOIS anti-harvesting is capable of reducing WHOIS-attributed spam email, which is the type of misuse with the highest frequency. And with that, I have concluded the presentation of the key findings of this WHOIS study and I will be happy - Nicolas and I will be happy to answer your questions.

Mary Wong: Thank you very much, Nektarios. And so if I may, Nathalie and the Operator to unmute all the attendees and participants from the phone bridge.

Coordinator: Thank you.

Mary Wong: Thank you. We will take questions and for those in the Adobe Connect Room, as I mentioned, you can type your question as a couple of people have already done or raise your hand, and for those on the audio bridge but not Adobe Connect, please just chime in and let us know if you have a question. Nektarios and Nicolas, I think we can start with Steve Metalitz's question in the chat and I'll read that out for those of you who are not in Adobe Connect.

It speaks to the survey portion of the study, specifically the registrant survey, where 57 responses were received but only 41 of those were complete. The

question therefore, and I think Nektarios and Nicolas, you addressed this somewhat in the report, but perhaps you can elaborate on it. How would the fact that only 41 out of 57 responses are complete affect the error rate?

Nektarios Leontiadis: Yes, so the numbers report associated only with the answers that are complete for each question. So, these in some way factors in the non-responses. So yes, we have higher - the area that is higher, and this is factored in based on the exclusion of the unanswered parts of the survey.

Mary Wong: Thanks, Nektarios. Steve, I don't know if you have a follow-up but I would also point to that part of the study and this is one reason or one useful aspect of having both the survey as the descriptive portion of the study and the experimental portion of the study so that some of the findings could be compared as I think Nektarios did earlier in the seminar. So hopefully that's helpful.

We have another question from (Dan Wright) and the question is whether the domain names -- and (Dan), I assume that you are referring to the experimental domains that were created and used by the researchers -- whether those domains resolve to working Web sites and whether you believe that that had any impact on the level of misuse.

Nektarios Leontiadis: Yes. So, there was no content associated with the experimental domains and that in fact is helpful in order to isolate the measured misuse from - not isolated but rather directly associated the misuse with WHOIS, and if we would add any content to the experimental domains, then we would probably be measuring the contribution of this content in the measured rate of misuse, and Nicolas wants to add something to that.

Nicolas Christin: Yes. So in addition to that another thing if you create some Web contents is that you may have some confounding factor in terms of how your advice is being publicized and so forth. Or you've got (unintelligible) search engine that finds your side or if, say, you're hosting it on shared infrastructure and

somebody else has your contact details, you cannot completely isolate where the misuse is coming from.

So, the design decision was very (right). It was definitely done to eliminate any potential confounding factors, and in that respect it may be viewed a little bit as a (unintelligible) bound on the type of problems that one can face, but at the same time we know we can confidently state that this is pure WHOIS misuse. We know that there isn't any other confounding factor that could explain why those domains were misused.

Mary Wong: Thank you, Nicolas. Thank you, Nektarios. (Dan) asked a couple of additional questions and (Dan), I'm going to ask your questions in two parts and following that we will deal with (Christina's) question and I see Kathy, you've got your hand up, and we'll take the queue that way.

So, one of (Dan's) other questions is, what were the forms anti-misuse or I suppose the countermeasures that were used by the registrars that the team used and which were the most effective?

Nektarios Leontiadis: Right. We - first of all this information is in the report and what I will say is that first we did not make any comparisons between the different types of - on the effectiveness of different types of anti-harvesting methods. We treated them collectively as one protective mechanism.

So there is no more detail on that, and in terms of what kind of protective measures that we're - in terms of the protective measures were deployed, we saw or we see the reports of rate limiting, blacklisting, IP blacklisting in a permanent or temporary fashion. We got also reports that registrars use (unintelligible) challenges and so on, and all this information is again in the report in terms of the frequency of use for each of these methods.

Mary Wong: Thank you, Nektarios. Just to add to that, I believe also that it would have been very difficult to separate the impact that each countermeasure would

have had for a specific or a particular domain name, and so no differentiation was done there as well, so it's a question of either you could see that there was misuse or you couldn't.

So this may be one of the questions that can be taken up in further analysis by the community as well. Then (Dan), your final question was whether any of the countermeasures or the methods that were used inhibit legitimate WHOIS queries from obtaining the information.

Nektarios Leontiadis: Is this question referring to the - it's not very clear to me as to who is the recipient or the person who's querying the WHOIS. We were not - in neither the experimental nor the descriptive part we were querying the WHOIS except from the test at the registrars and registries for the existence their harvesting measures. So, if you could clarify the question, it would be very helpful.

Mary Wong: Thank you. And (Dan), please feel free to follow up either in chat or orally. And while you do that, if you wish to do that, we will go on to Kristina Rosette's question, which is about the experimental study, and whether it collected or generated data that would allow identification or source of the person or the entity that was misusing the WHOIS data.

Nektarios Leontiadis: The only way this could be done from our perspective to, I mean, to identify WHOIS misusing the information, the WHOIS information, is by analyzing the center of the spam emails or by analyzing the phone numbers that are calling and they're leaving us spam voicemails. So, and Nicolas wants to add something.

Nicolas Christin: Yes, so you can analyze the centers. That's one thing that you can do. You can also typically those people are selling something, so you can try to analyze the type of products that they are selling and see if you can (unintelligible). This would not be very different from the analysis of large-scale spam (complaints) that were done in academic research.

We didn't do it, it was not within the charter of this study. Could it be done? Probably using the data that we had corrected, I don't know if we maintained it but this type of study could be done using pretty much the same mechanisms and just analyzing the existing data corrected as a result of these experiments.

Mary Wong: Thank you. And I should have added when I made the introduction that all of these findings and results of this and the other studies will be placed before the GNSO Council and the community for further review and analysis at the conclusion of this particular study. So, in terms of next steps, further work, and future work, some of these questions that are being raised might be useful to raise them again at that point.

Kathy, you're next.

Kathy Kleinman: Great. Excuse me, I have a cold. Can you hear me, Mary?

Mary Wong: Yes, go ahead.

Kathy Kleinman: Great. Thank you to the presenters. This is fascinating report and thank you for the presentation. A question, it was my sense in that there was some outreach to law enforcement and I was wondering if you could tell us a little bit about the response of law enforcement on the abuse of the WHOIS data, and then I have a follow-up question as well.

Nektarios Leontiadis: Thanks for your question. I don't go into detail into the specific, its specific response from law enforcement that we got as we are basically not allowed to individualize the responses but in essence they use WHOIS information either to identify or help them solve current cases or they also consider WHOIS data when analyzing attacks in order to see if it is a contributing factor to a case that they're working on.

Kathy Kleinman: Quick follow-up on that. Did you find that law enforcement was finding this to be a factor? For example, we found out in the US that our Department of Motor Vehicle records that showed people's home address have been used to stalk and kill an actress a number of years ago now. You know, are you finding things that have nothing to do with kind of email or phone violations but stalking, harassment, location of physical organizational entities or individuals' organizations or something like that. Did you find any of that?

Nektarios Leontiadis: I don't believe that we find something that, we found something that was at high occurrence. The responses that we did receive had some details on specific cases and we have tried to present them in the report in a way that they are not disclosing specifics. Nicolas (unintelligible)...

Kathy Kleinman: Great.

Nektarios Leontiadis: ...To that.

Nicolas Christin: Yes, so there's a balance to strike between preserving the anonymity of both the respondents and of their cases but they're referring to with informing the community. To go to your example of really egregious abuse, I wouldn't even talk about misuse at that point. We haven't seen any statistically significant reports from our law enforcement surveys of really egregious behavior coming from WHOIS misuse.

We did see, however, some indication that these things do happen and by these things I mean email spam, you know, phone number misuse, and things of that nature. This is something that without deanonymizing, you know, our responses, this is something that has been observed in a few instances. For more egregious types of crimes, I would say not to any measurable or statistically significant level.

Mary Wong: Thank you Kathy for the question and Nicolas and Nektarios. I should perhaps add that the law enforcement cybersecurity researchers and the first

responder study that was a part of a survey, that was the descriptive part of the study, and so it wasn't so much how these organizations like law enforcement would use the WHOIS data but it was more a question of what they would have experienced and so for example, the sort of incidents that would have been reported to them, and that was the scope of the survey and that was what was within the scope of this particular study.

Nektarios Leontiadis: That's correct.

Mary Wong: We have time for a few more questions. Holly Raiche, I don't know if you would like to ask a question. If so, whether you'd like to type it or speak orally. Please go ahead if you have a question, Holly. I see that Holly is typing. If anyone else has a question, please feel free to ask. Holly's question I see is, what were the most effective anti-harvesting mechanisms?

Nektarios, I think you addressed that somewhat in an earlier answer, but I wonder if you care to repeat and to elaborate at this point.

Nektarios Leontiadis: Yes, (unintelligible). Yes. As I said, we did not do a comparative analysis between the different types of anti-harvesting mechanisms deployed but we rather approach the existence of the anti-harvesting methods as one thing. We didn't go into specifics as to the effect of each of those measures.

Mary Wong: Thank you. And we have another question from Kathy, and the question is whether there was any outreach to attorneys, particularly those who work frequently with registrants, for the type of abuse or misuse of the WHOIS data that they might see, including for what purposes. While Nektarios come back on the line and Nicolas, too.

Kathy, I guess the attorneys not necessarily as a group might have been surveyed as part of the registrant survey. I don't believe that they were included or that they were a group in any of the other categories that were surveyed. Nektarios, Nicolas, is that correct?

Nektarios Leontiadis: Yes. Thank you, Mary. Yes. That is correct. We did not in the registrant sample, in the registrant survey, I'm sorry, we did not target specifically any group or type of company for participation in the survey. For the law enforcement security researcher survey, we mainly targeted invited people that would be able to contribute in giving us a more - greater perspective on the occurrence of misuse.

Now on the - whether or not we had this kind of attorneys invited at these survey, I think I'll have to defer and double take or invite (unintelligible).

Mary Wong: Thank you. We have time for a few more questions, and actually at this point I'm going to call on Lisa Pfeiffer, a consultant who has been with us on these studies, including this one, to ask Lisa if you have anything to add to either this or any of the other questions and discussions that we've been having.

Lisa Pfeiffer: Thanks, Mary. I just wanted to follow up on Nektarios' point. I actually had occasion to go back and look at the categories of people that were surveyed in that expert survey, and we did discuss including attorneys that might investigate, for example, IP theft, as part of that survey and the decision was made to let that happen more organically as part of the registrant survey.

And Nektarios, you might want to comment a little bit about the composition of the responses you got from the registrant survey. I believe about half of the responses came from self-identified businesses or registrants that had over 10 domains.

Nektarios Leontiadis: Yes, that is correct. And in fact about half of the responses were from for-profit organizations. About 30% were individuals, about 14% were nonprofit organizations, and about 5% were informal interest groups.

Mary Wong: Thanks, Lisa. Thanks, Nektarios. In lieu of some of the questions and I believe that there may be a couple more coming in either through the chat

room or by the audio bridge, I was wondering, Nicolas and Nektarios, if you would care to talk a little bit about some of the challenges that you face in this study because I think some of the questions probably relate directly to some of those challenges, and perhaps particularly some of the limitations that you found in doing these studies relating to whether or not are the kinds of misuse that could or could not be usefully studied at least at a statistically significant level.

Nektarios Leontiadis: Sure. In terms of the methodology, the greatest difficulty I think was with the registrant survey as even though we made numbers translations for the survey, we offered prizes for the completion of the survey. We ended up having a very small turnaround and that was very unfortunate. Now, in terms of the registrar and registry survey, we also found resistance in getting responses from those parties and we believe that this was mainly because the registrars and registries did not feel any obligation to serve this kind of data.

And so we did not manage to get the coverage that we would want. Now in terms of the kind of things that cannot be adequately measured in terms of WHOIS misuse, as that's for example identity theft, let's say. This, something like that would need to - a measurement that would be able to measure that would have to go on a much greater - for much greater time so six months in the case that of our experiment would not be able to adequately capture such occurrence.

And also, the way that someone would be able to observe the occurrence of identity theft is also something that cannot be easily be designed as part of an experiment.

Mary Wong: Thank you. I know that there a couple of more questions that have come in, one from Steve Metalitz on the numbers. And Steve, I believe that you're referring back to the registrant survey where 41 out of 57 responses were completed, and your question is whether that means it was 12 individual

registrants who responded. Nektarios, I don't know if you have those numbers handy.

Nektarios Leontiadis: I don't. No, actually I don't have these. I mean these, what Steve says makes sense but I have to double check this with our data.

Mary Wong: Right. And so Steve, I would say it's in the report. We can certainly double check on that and get back to you with the specific answer after this webinar. And Kristina, you had a follow-up question. If an IP attorney (unintelligible) someone else called a registrant or sent an email to an registrant of an infringing domain, and if that registrant was part of the survey that was done and then reported either that call or that email is misused, would that type of contact from the attorney then be considered misuse for purposes of the survey or is it filtered out?

Nektarios, I guess you can give a full answer to this. I believe that the survey was conducted and designed to be self-reporting.

Nektarios Leontiadis: Let me read this question again. I think that's the problem with identifying WHOIS-attributed email misuse on the part of the registrants. And as I said earlier in the presentation this difficulty might be contributing to the difference in the measured versus the reported rate of email misuse. Now this is after the registrant on whether he or she will identify this as a WHOIS-attributed misuse or not. So, I don't have a definite answer on that from the perspective of the registrant.

Now from the perspective of the experimental study, these would be considered as WHOIS-attributed misuse, and that is because of the definition we used as to what constitutes spam email, and that is when an email is unsolicited and the recipient has not provided explicitly a concern to receipt of an email. So this qualifies as spam email.

Mary Wong: And I should add that the terms and the definitions that we used by our research team in this study as well as in all the other studies were consistent to the extent that the same terms were used across the different studies and they're also included either in the terms of reference or repeated in the study results or in both cases.

I think we can take one, possibly two depending on the nature of the discussion, further questions, and from one (Ahmer). The study shows a confidence rate of 95%. To what extent do you believe the statistical significance of the results is generalizable? At least amongst those five GTLDs that were included.

Nektarios Leontiadis: So the meaning of these statistical parameters essentially tells us that for 95% of the population of the domains in those DLDs are reported values or findings will essentially be what you see in 12.7% of the actual values in the whole population. So, they are generalizable within this extent. Now, for the remaining 5% of the population, this distance from the reported values, the measured values, can be greater or lower than 12.7%, which is our error rate.

Mary Wong: Thank you. And perhaps we can take that last question that was just typed into the Adobe Connect Room by Kathy as to the type of email spam that was observed and I believe that part of this may have been answered by Nektarios' pointing us to the definition of spam that was used for purposes of this study, but Nektarios, please feel free to go ahead and elaborate.

Nicolas Christin: Well actually this is Nicolas. I'm going to take this one. The type of email spam we did see was I think, and I would actually need - we would actually need to dig up the data corpus if we haven't talked about it in the report but I think mostly it was from Web-based services, search engine optimization type of things, (unintelligible) more spam like (unintelligible) spam but mostly I would say somewhat advertisements for domain owners that was, I think, the bulk of it.

Mary Wong: Thank you, Nicolas. That is very helpful. And again, I should note for everyone's benefit that the study, in fact, none of those studies was about the content either of Web sites or emails, and so hopefully that is helpful as well. We're actually now at the top of the hour and so it probably behooves me to thank our presenters first of all for the research and the analysis that they have done and I see on some of the comments in the chat that all that work was definitely very much appreciated.

And I will remind everyone, then, that the public comment forum period is still open. The link to both the background as well as to where you can download the study itself is here on this slide, and the comment forum is actually open through 18 of January, even though we designated initial public comment and a reply period.

We will be accepting comments through the 18 of January following which those results will be summarized and analyzed in a report put out, and a final report then from the team to go to the GNSO for its consideration. So, thank you, Nicolas. Thank you, Nektarios. Thank you to everybody for attending, we will post these slides, the transcript, and the recording as soon as possible.

Man: Thank you, Mary.

Mary Wong: Thank you very much.

Man: Thank you.

Mary Wong: And Operator, we can now stop the recording. Thank you.

END